Hash Property and Fixed-rate Universal Coding Theorems

Jun Muramatsu *Member, IEEE,* Shigeki Miyake *Member, IEEE,*

Abstract

The aim of this paper is to prove the achievability of fixed-rate universal coding problems by using our previously introduced notion of hash property. These problems are the fixed-rate lossless universal source coding problem and the fixed-rate universal channel coding problem. Since an ensemble of sparse matrices satisfies the hash property requirement, it is proved that we can construct universal codes by using sparse matrices.

Index Terms

channel coding, fixed-rate universal codes hash functions, linear codes, lossless source coding, minimum-divergence encoding, minimum-entropy decoding, shannon theory, sparse matrix

I. Introduction

The notion of hash property is introduced in [12]. It is a sufficient condition for the achievability of coding theorems including lossless and lossy source coding, channel coding, the Slepian-Wolf problem, the Wyner-Ziv problem, the Gel'fand-Pinsker problem, and the problem of source coding with partial side information at the decoder. Since an ensemble of sparse matrices satisfies the hash property requirement, it is proved that we can construct codes by using sparse matrices and maximum-likelihood coding.

However, it is assumed in [12] that source and channel distributions are used when designing a code. The aim of this paper is to prove fixed-rate universal coding theorems based on the hash property, where a specific probability distribution is not assumed for the design of a code and the error probability of a code vanishes for all sources specified by the encoding rate.

We prove theorems of fixed-rate lossless universal source coding (see Fig. 1) and fixed-rate universal channel coding (see Fig. 2). In the construction of codes, the maximum-likelihood coding used in [12] is replaced by a minimum-divergence encoder and a minimum-entropy decoder. A practical algorithm has been obtained for the minimum-entropy decoder by using linear programming [2]. It should be noted that a practical algorithm for the minimum-divergence encoder can also be obtained by using linear programming as shown in Section V. The fixed-rate lossless universal source coding theorem is proved in [3] for the ensemble of all linear matrices in the context of the Slepian-Wolf source coding problem, in [7] for the class of universal hash functions, and

J. Muramatsu is with NTT Communication Science Laboratories, NTT Corporation, 2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan (E-mail: pure@cslab.kecl.ntt.co.jp). S. Miyake is with NTT Network Innovation Laboratories, NTT Corporation, 1-1, Hikarinooka, Yokosuka-shi, Kanagawa 239-0847, Japan (E-mail: miyake.shigeki@lab.ntt.co.jp).

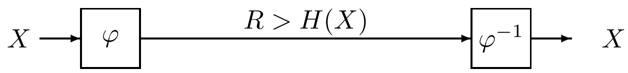

Fig. 1. Lossless Source Coding

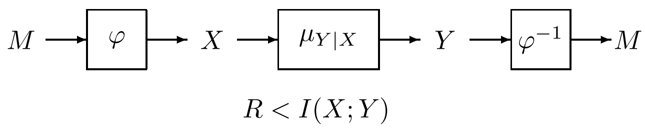

Fig. 2. Channel Coding

in [11] implicitly for an ensemble of sparse matrices in the context of a secret key agreement from correlated source outputs. The universal channel coding theorem that employs sparse matrices is proved in [8] for an additive noise channel and in [9] for an arbitrary channel. It should be noted here that the linearity for an ensemble member is not assumed in our proof. Our proof assumes that ensembles of sparse matrices have a hash property and so is simpler than previously reported proofs [11][8][9].

II. Definitions and Notations

Throughout this paper, we use the following definitions and notations.

Column vectors and sequences are denoted in boldface. Let $A\boldsymbol{u}$ denote a value taken by a function $A : \mathcal{U}^n \to \overline{\mathcal{U}}$ at $\boldsymbol{u} \in \mathcal{U}^n$ where $\mathcal{U}^n$ is a domain of the function. It should be noted that A may be non-linear. For a function A and a set of functions $\mathcal{A}$, let $\mathrm{Im}A$ and $\mathrm{Im}\mathcal{A}$ be defined as

$$\mathrm{Im}A \equiv \{A\boldsymbol{u} : \boldsymbol{u} \in \mathcal{U}^n\}$$

$$\mathrm{Im}\mathcal{A} \equiv \bigcup_{A \in \mathcal{A}} \mathrm{Im}A.$$

The cardinality of a set $\mathcal{U}$ is denoted by $|\mathcal{U}|$ and $\mathcal{U} - \{\boldsymbol{u}\}$ is a set difference. We define sets $\mathcal{C}_A(\boldsymbol{c})$ and $\mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m})$ as

$$\mathcal{C}_A(\boldsymbol{c}) \equiv \{\boldsymbol{u} : A\boldsymbol{u} = \boldsymbol{c}\}$$

$$\mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m}) \equiv \{\boldsymbol{u} : A\boldsymbol{u} = \boldsymbol{c}, B\boldsymbol{u} = \boldsymbol{m}\}.$$

In the context of linear codes, $\mathcal{C}_A(\boldsymbol{c})$ is called a coset determined by $\boldsymbol{c}$.

Let p and p' be probability distributions and let q and q' be conditional probability distributions. Then entropy $H(p)$, conditional entropy $H(q|p)$, divergence $D(p\|p')$, and conditional divergence $D(q\|q'|p)$ are defined as

$$H(p) \equiv \sum_u p(u) \log \frac{1}{p(u)}$$

$$H(q|p) \equiv \sum_{u,v} q(u|v)p(v) \log \frac{1}{q(u|v)}$$

$$D(p \parallel p') \equiv \sum_u p(u) \log \frac{p(u)}{p'(u)}$$

$$D(q \parallel q'|p) \equiv \sum_{v} p(v) \sum_{u} q(u|v) \log \frac{q(u|v)}{q'(u|v)},$$

where we assume the base 2 of the logarithm.

Let μ_{UV} be the joint probability distribution of random variables U and V. Let μ_U and μ_V be the respective marginal distributions and $\mu_{U|V}$ be the conditional probability distribution. Then the entropy $H(U)$, the conditional entropy $H(U|V)$, and the mutual information $I(U;V)$ of random variables are defined as

$$\begin{aligned} H(U) &\equiv H(\mu_U) \\ H(U|V) &\equiv H(\mu_{U|V}|\mu_V) \\ I(U;V) &\equiv H(\mu_U) + H(\mu_V) - H(\mu_{UV}). \end{aligned}$$

Let $\nu_{\boldsymbol{u}}$ and $\nu_{\boldsymbol{u}|\boldsymbol{v}}$ be defined as

$$\begin{aligned} \nu_{\boldsymbol{u}}(u) &\equiv \frac{|\{1 \leq i \leq n : u_i = u\}|}{n} \\ \nu_{\boldsymbol{u}|\boldsymbol{v}}(u|v) &\equiv \frac{\nu_{\boldsymbol{u}\boldsymbol{v}}(u,v)}{\nu_{\boldsymbol{v}}(v)}. \end{aligned}$$

We call $\nu_{\boldsymbol{u}}$ a type [1] of $\boldsymbol{u} \in \mathcal{U}^n$ and $\nu_{\boldsymbol{u}|\boldsymbol{v}}$ a conditional type. Let $U \equiv \nu_U$ be the type of a sequence and $U|V \equiv \nu_{U|V}$ be the conditional type of a sequence given a sequence of type U. Then a set of typical sequences $\mathcal{T}_U$ and a set of conditionally typical sequences $\mathcal{T}_{U|V}(\boldsymbol{v})$ are defined as

$$\begin{aligned} \mathcal{T}_U &\equiv \{\boldsymbol{u} : \nu_{\boldsymbol{u}} = \nu_U\} \\ \mathcal{T}_{U|V}(\boldsymbol{v}) &\equiv \left\{\boldsymbol{u} : \nu_{\boldsymbol{u}|\boldsymbol{v}} = \nu_{U|V}\right\}, \end{aligned}$$

respectively. The empirical entropy, the empirical conditional entropy, and empirical mutual information are defined as

$$\begin{aligned} H(\boldsymbol{u}) &\equiv H(\nu_{\boldsymbol{u}}) \\ H(\boldsymbol{u}|\boldsymbol{v}) &\equiv H(\nu_{\boldsymbol{u}|\boldsymbol{v}}|\nu_{\boldsymbol{v}}) \\ I(\boldsymbol{u};\boldsymbol{v}) &\equiv H(\nu_{\boldsymbol{u}}) + H(\nu_{\boldsymbol{v}}) - H(\nu_{\boldsymbol{u}\boldsymbol{v}}). \end{aligned}$$

In the construction of a universal source code, we use a *minimum-entropy decoder*

$$g_A(\boldsymbol{c}) \equiv \arg \min_{\boldsymbol{x}' \in \mathcal{C}_A(\boldsymbol{c})} H(\boldsymbol{x}')$$

It should be noted that the linear programing technique introduced in [2] can be applied to the minimum-entropy decoder g_A. In the construction of a universal channel code, we use a *minimum-divergence encoder*

$$g_{AB}(\boldsymbol{c},\boldsymbol{m}) \equiv \arg \min_{\boldsymbol{x}' \in \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})} D(\nu_{\boldsymbol{x}'} \| \mu_X)$$

and a minimum-entropy decoder

$$g_A(\boldsymbol{c},\boldsymbol{y}) \equiv \arg \min_{\boldsymbol{x}' \in \mathcal{C}_A(\boldsymbol{c})} H(\boldsymbol{x}'|\boldsymbol{y}).$$

[1] In [12], the type of a sequence is defined as a histogram $\{n\nu_{\boldsymbol{u}}(u)\}_{u \in \mathcal{U}}$.

It should be noted that we have

$$\begin{aligned} g_{AB}(\boldsymbol{c},\boldsymbol{m}) &= \arg\max_{\boldsymbol{x}'\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})} [\log\mu_X(\boldsymbol{x}') + nH(\nu_{\boldsymbol{x}'})] \\ &= \arg\max_{U'} \left[nH(U') + \max_{\boldsymbol{x}'\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})\cap\mathcal{T}_{U'}} \log\mu_X(\boldsymbol{x}') \right] \end{aligned}$$

from Lemma 7. When functions A and B are linear, the linear programing technieque introduced in [6] can be applied to the maximization $\max_{\boldsymbol{x}'}\mu_X(\boldsymbol{x}')$ because U' is fixed and the constraint condition $\boldsymbol{x}'\in\mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m})\cap\mathcal{T}_{U'}$ is represented by linear functions.

Finally, we define $\chi(\cdot)$ as

$$\chi(a=b) \equiv \begin{cases} 1, & \text{if } a=b \\ 0, & \text{if } a\neq b \end{cases}$$

$$\chi(a\neq b) \equiv \begin{cases} 1, & \text{if } a\neq b \\ 0, & \text{if } a=b. \end{cases}$$

We define a sequence $\{\lambda_{\mathcal{U}}(n)\}_{n=1}^{\infty}$ as

$$\lambda_{\mathcal{U}}(n) \equiv \frac{|\mathcal{U}|\log[n+1]}{n}. \tag{1}$$

It should be noted here that the product set $\mathcal{U}\times\mathcal{V}$ is denoted by $\mathcal{U}\mathcal{V}$ when it appears in the subscript of this function and we omit argument n of $\lambda_{\mathcal{U}}$ when n is clear in the context. We define $|\cdot|^+$ as

$$|\theta|^+ \equiv \begin{cases} \theta, & \text{if } \theta>0, \\ 0, & \text{if } \theta\leq 0. \end{cases} \tag{2}$$

III. $(\boldsymbol{\alpha},\boldsymbol{\beta})$-HASH PROPERTY

In this section, we reveiw the notion of the $(\boldsymbol{\alpha},\boldsymbol{\beta})$-hash property introduced in [12]. This is a sufficient condition for coding theorems, where the linearity of functions is not assumed. By using this notion, we prove a fixed-rate universal source coding theorem and a fixed-rate universal source coding theorem.

Throughout the paper, $A\boldsymbol{u}$ denotes a value taken by a function A at $\boldsymbol{u}\in\mathcal{U}^n$ where $\mathcal{U}^n$ is the domain of the function. It should again be noted here that A may be non-linear. We define the $(\boldsymbol{\alpha},\boldsymbol{\beta})$-hash property in the following.

Definition 1: Let $\mathcal{A}$ be a set of functions $A:\mathcal{U}^n\to\overline{\mathcal{U}}$ and we assume that $\mathrm{Im}A=\mathrm{Im}\mathcal{A}$ for all $A\in\mathcal{A}$ and

$$\lim_{n\to\infty}\frac{\log\frac{|\overline{\mathcal{U}}|}{|\mathrm{Im}\mathcal{A}|}}{n} = 0. \tag{H1}$$

Let p_A be a probability distribution on $\mathcal{A}$. We call a pair $(\mathcal{A},p_A)$ an *ensemble*. Then, $(\mathcal{A},p_A)$ has an $(\boldsymbol{\alpha},\boldsymbol{\beta})$*-hash property* if $\boldsymbol{\alpha}\equiv\{\alpha(n)\}_{n=1}^{\infty}$ and $\boldsymbol{\beta}\equiv\{\beta(n)\}_{n=1}^{\infty}$ satisfy

$$\lim_{n\to\infty}\alpha(n) = 1 \tag{H2}$$

$$\lim_{n\to\infty}\beta(n) = 0 \tag{H3}$$

and

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{T}\\ \boldsymbol{u}'\in\mathcal{T}'}} p\left(\{A : A\boldsymbol{u} = A\boldsymbol{u}'\}\right) \leq |\mathcal{T}\cap\mathcal{T}'| + \frac{|\mathcal{T}||\mathcal{T}'|\alpha(n)}{|\mathrm{Im}\mathcal{A}|} + \min\{|\mathcal{T}|,|\mathcal{T}'|\}\beta(n) \tag{H4}$$

for any $\mathcal{T},\mathcal{T}'\subset\mathcal{U}^n$. Throughout this paper, we omit argument n of α and β when n is fixed. ■

In the following, we present two examples of ensembles that have a hash property.

Example 1: In this example, we consider a universal class of hash functions introduced in [5]. A set $\mathcal{A}$ of functions $A:\mathcal{U}^n\to\overline{\mathcal{U}}$ is called a *universal class of hash functions* if

$$|\{A : A\boldsymbol{u} = A\boldsymbol{u}'\}| \leq \frac{|\mathcal{A}|}{|\overline{\mathcal{U}}|}$$

for any $\boldsymbol{u}\neq\boldsymbol{u}'$. For example, the set of all functions on $\mathcal{U}^n$ and the set of all linear functions $A:\mathcal{U}^n\to\mathcal{U}^{l_A}$ are universal classes of hash functions (see [5]).

It should be noted that every example above satisfies $\mathrm{Im}\mathcal{A}=\overline{\mathcal{U}}$. When $\mathcal{A}$ is a universal class of hash functions and p_A is the uniform probability on $\mathcal{A}$, we have

$$\sum_{\substack{\boldsymbol{u}\in\mathcal{T}\\ \boldsymbol{u}'\in\mathcal{T}'}} p_A\left(\{A : A\boldsymbol{u} = A\boldsymbol{u}'\}\right) \leq |\mathcal{T}\cap\mathcal{T}'| + \frac{|\mathcal{T}||\mathcal{T}'|}{|\mathrm{Im}\mathcal{A}|}.$$

This implies that $(\mathcal{A},p_A)$ has a $(\mathbf{1},\mathbf{0})$-hash property, where $\alpha(n)\equiv 1$ and $\beta(n)\equiv 0$ for every n. ■

Example 2: In this example, we revew the ensemble of q-ary sparse matrices introduced in [12]. In the following, let $\mathcal{U}\equiv\mathrm{GF}(q)$ and $l_A\equiv nR$. We generate an $l\times n$ matrix A with the following procedure:

1) Start from an all-zero matrix.
2) For each $i\in\{1,\ldots,n\}$, repeat the following procedure τ times:
   a) Choose $(j,a)\in\{1,\ldots,l_A\}\times[\mathrm{GF}(q)-\{0\}]$ uniformly at random.
   b) Add a to the (j,i) component of A.

Let $(\mathcal{A},p_A)$ be an ensemble corresponding to the above procedure. Then

$$\mathrm{Im}A = \begin{cases} \left\{\boldsymbol{u}\in\mathcal{U}^l : \begin{array}{l}\boldsymbol{u}\text{ has an even number of}\\ \text{non-zero elements}\end{array}\right\}, & \text{if } q=2\\ \mathcal{U}^l, & \text{if } q>2\end{cases}$$

for all $A\in\mathcal{A}$ and there is $(\boldsymbol{\alpha}_A,\boldsymbol{\beta}_A)$ such that $(\mathcal{A},p_A)$ has an $(\boldsymbol{\alpha}_A,\boldsymbol{\beta}_A)$-hash property (see [12, Theorem 2]). ■

In the following, Let $\mathcal{A}$ (resp. $\mathcal{B}$) be a set of functions $A:\mathcal{U}^n\to\overline{\mathcal{U}}_A$ (resp. $B:\mathcal{U}^n\to\overline{\mathcal{U}}_B$). We assume that an ensemble $(\mathcal{A},p_A)$ has an $(\boldsymbol{\alpha}_A,\boldsymbol{\beta}_A)$-hash property and an ensemble $(\mathcal{A}\times\mathcal{B},p_A\times p_B)$ also has an $(\boldsymbol{\alpha}_{AB},\boldsymbol{\beta}_{AB})$-hash property. We also assume that p_C and p_M is the uniform distribution on $\mathrm{Im}\mathcal{A}$ and $\mathrm{Im}\mathcal{B}$, respectively, and random variables A, B, C, and M are mutually independent, that is,

$$p_C(\boldsymbol{c}) = \begin{cases}\frac{1}{|\mathrm{Im}\mathcal{A}|}, & \text{if } \boldsymbol{c}\in\mathrm{Im}\mathcal{A}\\ 0, & \text{if } \boldsymbol{c}\in\overline{\mathcal{U}}-\mathrm{Im}\mathcal{A}\end{cases}$$

$$p_M(\boldsymbol{m}) = \begin{cases}\frac{1}{|\mathrm{Im}\mathcal{B}|}, & \text{if } \boldsymbol{m}\in\mathrm{Im}\mathcal{B}\\ 0, & \text{if } \boldsymbol{m}\in\overline{\mathcal{U}}-\mathrm{Im}\mathcal{A}\end{cases}$$

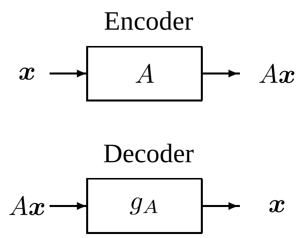

Fig. 3. Construction of Fixed-rate Source Code

$$p_{ABCM}(A, B, \boldsymbol{c}, \boldsymbol{m}) = p_A(A)p_B(B)p_C(\boldsymbol{c})p_M(\boldsymbol{m})$$

for any A, B, and $\boldsymbol{c}$. We use the following lemmas, which are shown in [12].

Lemma 1 ([12, Lemma 9]): For any A and $\boldsymbol{u} \in \mathcal{U}^n$,

$$p_C(\{\boldsymbol{c} : A\boldsymbol{u} = \boldsymbol{c}\}) = \sum_{\boldsymbol{c}} p_C(\boldsymbol{c})\chi(A\boldsymbol{u} = \boldsymbol{c}) = \frac{1}{|\mathrm{Im}\mathcal{A}|}$$

and for any $\boldsymbol{u} \in \mathcal{U}^n$,

$$E_{AC}[\chi(A\boldsymbol{u} = \boldsymbol{c})] = \sum_{A,\boldsymbol{c}} p_{AC}(A, \boldsymbol{c})\chi(A\boldsymbol{u} = \boldsymbol{c}) = \frac{1}{|\mathrm{Im}\mathcal{A}|}.$$

Lemma 2 ([12, Lemma 2]): If $\mathcal{G} \subset \mathcal{U}^n$ and $\boldsymbol{u} \notin \mathcal{G}$, then

$$p_A(\{A : \mathcal{G} \cap \mathcal{C}_A(A\boldsymbol{u}) \neq \emptyset\}) \leq \frac{|\mathcal{G}|\alpha_A}{|\mathrm{Im}\mathcal{A}|} + \beta_A.$$

■

Lemma 3 ([12, Lemma 5]): If $\mathcal{T} \neq \emptyset$, then

$$p_{ABCM}(\{(A, B, \boldsymbol{c}, \boldsymbol{m}) : \mathcal{T} \cap \mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m}) = \emptyset\}) \leq \alpha_{AB} - 1 + \frac{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|\left[\beta_{AB} + 1\right]}{|\mathcal{T}|}.$$

■

When $(\mathcal{A}, p_A)$ and $(\mathcal{B}, p_B)$ are the ensembles of $l_A \times n$ and $l_B \times n$ linear matrices, respectively, we have the following lemma.

Lemma 4 ([12, Lemma 7]): The joint distribution $(\mathcal{A} \times \mathcal{B}, p_{AB})$ has an $(\boldsymbol{\alpha}_{AB}, \boldsymbol{\beta}_{AB})$-hash property for the ensemble of functions $A \oplus B : \mathcal{U}^n \to \mathcal{U}^{l_A + l_B}$ defined as

$$A \oplus B(\boldsymbol{u}) \equiv (A\boldsymbol{u}, B\boldsymbol{u}),$$

where

$$\alpha_{AB}(n) = \alpha_A(n)\alpha_B(n) \tag{3}$$

$$\beta_{AB}(n) = \min\{\beta_A(n), \beta_B(n)\}. \tag{4}$$

■

IV. Fixed-rate Lossless Universal Source Coding

In this section, we consider the fixed-rate lossless universal source coding illustrated in Fig. 1.

For a given encoding rate R, l_A is given by

$$l_A \equiv \frac{nR}{\log|\mathcal{X}|}.$$

We fix a function

$$A : \mathcal{X}^n \to \mathcal{X}^{l_A}$$

which is available to construct an encoder and a decoder. We define the encoder and the decoder (illustrated in Fig. 3)

$$\varphi_X : \mathcal{X}^n \to \mathcal{X}^{l_A}$$
$$\varphi^{-1} : \mathcal{X}^{l_A} \to \mathcal{X}^n$$

as

$$\varphi(\boldsymbol{x}) \equiv A\boldsymbol{x}$$
$$\varphi^{-1}(\boldsymbol{c}) \equiv g_A(\boldsymbol{c}),$$

where

$$g_A(\boldsymbol{c}) \equiv \arg\min_{\boldsymbol{x}' \in \mathcal{C}_A(\boldsymbol{c})} H(\boldsymbol{x}').$$

The error probability $\mathrm{Error}_X(A)$ is given by

$$\mathrm{Error}_X(A) \equiv \mu_X\left(\left\{\boldsymbol{x} : \varphi^{-1}(\varphi(\boldsymbol{x})) \neq \boldsymbol{x}\right\}\right).$$

We have the following theorem. It should be noted that the alphabet $\mathcal{X}$ may not be binary.

Theorem 1: Assume that an ensemble $(\mathcal{A}, p_A)$ has an $(\boldsymbol{\alpha}_A, \boldsymbol{\beta}_A)$-hash property. For a fixed rate R, $\delta > 0$ and a sufficiently large n, there is a function (matrix) $A \in \mathcal{A}$ such that

$$\mathrm{Error}_X(A) \leq \max\left\{\frac{\alpha_A|\mathcal{X}|^{l_A}}{|\mathrm{Im}\mathcal{A}|}, 1\right\} 2^{-n[\inf F_X(R) - 2\lambda_{\mathcal{X}}]} + \beta_A \tag{5}$$

for any stationary memoryless sources X satisfying

$$H(X) < R, \tag{6}$$

where

$$F_X(R) \equiv \min_{U'}\left[D(\nu_{U'}\|\mu_X) + |R - H(U')|^+\right]$$

and the infimum is taken over all X satisfying (6). Since

$$\inf_{X : H(X) > R} F_X(R) > 0,$$

then the error probability goes to zero as $n \to \infty$ for all X satisfying (6). ∎

We can prove the coding theorem for a channel $\mu_{Y|X}$ with additive noise $Z \equiv Y - X$ by letting A and $\mathcal{C}_A(\mathbf{0}) = \{\boldsymbol{x} : A\boldsymbol{x} = \mathbf{0}\}$ be a parity check matrix and a set of codewords (channel inputs), respectively. Then the encoding rate of this channel code is given by

$$\frac{\log |\mathcal{C}_A(\mathbf{0})|}{n} \geq \log |\mathcal{X}| - R$$

and the error probability is given as

$$\mathrm{Error}_{Y|X}(A) \equiv \frac{1}{|\mathcal{C}_A(\mathbf{0})|} \sum_{\boldsymbol{x} \in \mathcal{C}_A(\mathbf{0})} \mu_{Y|X}\left(\{\boldsymbol{y} : g_A(A\boldsymbol{y}) \neq \boldsymbol{y} - \boldsymbol{x}\} \mid \boldsymbol{x}\right).$$

Since

$$\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{x}$$

$$A\boldsymbol{z} = A\boldsymbol{y} - A\boldsymbol{x} = A\boldsymbol{y},$$

then the decoding of channel input $\boldsymbol{x}$ from a syndrome $A\boldsymbol{y}$ is equivalent to the decoding of source output $\boldsymbol{z}$ from its codeword $A\boldsymbol{z}$ by using g_A. We have the following corollary.

Corollary 2: Assume that an ensemble $(\mathcal{A}, p_A)$ of linear functions has an $(\boldsymbol{\alpha}_A, \boldsymbol{\beta}_A)$-hash property. For a fixed rate R, $\delta > 0$ and sufficiently large n, there is a (sparse) matrix $A \in \mathcal{A}$ such that

$$\mathrm{Error}_{Y|X}(A) \leq \max\left\{\frac{\alpha_A |\mathcal{X}|^{l_A}}{|\mathrm{Im}\mathcal{A}|}, 1\right\} 2^{-n[\inf F_Z(R) - 2\lambda_{\mathcal{X}}]} + \beta_A$$

for any stationary memoryless channel with additive noize Z satisfying

$$\log |\mathcal{X}| - R < I(X;Y) = \log |\mathcal{X}| - H(Z), \tag{7}$$

where the infimum is taken over all Z satisfying (7) and the error probability goes to zero as $n \to \infty$ for all X satisfying (7). ■

Remark 1: It should be noted here that the condition (H2) can be replaced by

$$\lim_{n \to \infty} \frac{\log \alpha_A(n)}{n} = 0. \tag{8}$$

By using the expurgation technique described in [1], we obtain an ensemble of sparce matrices that have an $(\boldsymbol{\alpha}_A, \mathbf{0})$-hash property, where (H2) is replaced by (8). This implies that we can omit the term β_A from the upper bound of the error probability. ■

Remark 2: Since a class of universal hash functions with a uniform distribution and an ensemble of all linear functions has a $(\mathbf{1}, \mathbf{0})$-hash property, we obtain the same results as those reported in [7] and [3], respectively, where F_X represents the error exponent function. When $(\mathcal{A}, p_A)$ is an ensemble of sparse matrices and $(\boldsymbol{\alpha}_A, \boldsymbol{\beta}_A)$ is defined properly, we have the same result as that found in [8]. ■

V. Fixed-rate Universal Channel Coding

The code for the channel coding problem (illustrated in Fig. 2) is given in the following (illustrated in Fig. 4). The idea for the construction is drawn from [10][12][9]. We give the explicit construction of the encoder by using minimum-divergence encoding, which is not described in [10][12][9].

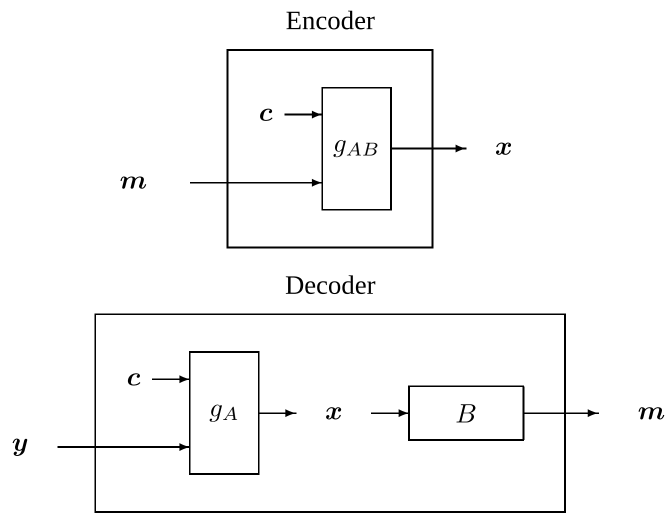

Fig. 4. Construction of Channel Code

For a given $R_A, R_B > 0$, let

$$A : \mathcal{X}^n \to \mathcal{X}^{l_A}$$
$$B : \mathcal{X}^n \to \mathcal{X}^{l_B}$$

satisfying

$$R_A = \frac{\log |\mathrm{Im} A|}{n}$$
$$R_B = \frac{\log |\mathrm{Im} B|}{n},$$

respectively.

We fix functions A, B and a vector $\boldsymbol{c}_n \in \mathcal{X}^{l_A}$ available to constract an encoder and a decoder.

We define the encoder and the decoder

$$\varphi : \mathcal{X}^{l_B} \to \mathcal{X}^n$$
$$\varphi^{-1} : \mathcal{Y}^n \to \mathcal{X}^{l_B}$$

as

$$\varphi(\boldsymbol{m}) \equiv g_{AB}(\boldsymbol{c}, \boldsymbol{m})$$
$$\varphi^{-1}(\boldsymbol{y}) \equiv B g_A(\boldsymbol{c}, \boldsymbol{y}),$$

where

$$g_{AB}(\boldsymbol{c}, \boldsymbol{m}) \equiv \arg \min_{\boldsymbol{x}' \in \mathcal{C}_{AB}(\boldsymbol{c}, \boldsymbol{m})} D(\nu_{\boldsymbol{x}'} \| \mu_X)$$
$$g_A(\boldsymbol{c}, \boldsymbol{y}) \equiv \arg \min_{\boldsymbol{x}' \in \mathcal{C}_A(\boldsymbol{c})} H(\boldsymbol{x}' | \boldsymbol{y}).$$

The error probability $\mathrm{Error}_{Y|X}(A,B,\boldsymbol{c})$ is given by

$$\mathrm{Error}_{Y|X}(A,B,\boldsymbol{c}) \equiv \sum_{\boldsymbol{m},\boldsymbol{y}} p_M(\boldsymbol{m})\mu_{Y|X}(\boldsymbol{y}|\varphi(\boldsymbol{m}))\chi(\varphi^{-1}(\boldsymbol{y}) \neq \boldsymbol{m}),$$

where

$$p_M(\boldsymbol{m}) \equiv \begin{cases} \frac{1}{|\mathrm{Im}B|}, & \text{if } \boldsymbol{c} \in \mathrm{Im}B \\ 0 & \text{if } \boldsymbol{c} \notin \mathrm{Im}B. \end{cases}$$

It should be noted that $\mathrm{Im}B$ represents a set of all messages and R_B represents the encoding rate of a channel.

We have the following theorem.

Theorem 3: Assume that an ensemble $(\mathcal{A}, p_A)$ (resp. $(\mathcal{A}\times\mathcal{B}, p_{AB})$) has an $(\boldsymbol{\alpha}_A, \boldsymbol{\beta}_A)$-hash (resp. $(\boldsymbol{\alpha}_{AB}, \boldsymbol{\beta}_{AB})$-hash) property. For a fixed rate $R_A, R_B > 0$, a given input distribution μ_X satisfying

$$H(X) > R_A + R_B, \tag{9}$$

$\delta > 0$, and a sufficiently large n, there are functions (matrices) $A \in \mathcal{A}$, $B \in \mathcal{B}$, and a vector $\boldsymbol{c} \in \mathrm{Im}A$ such that

$$\mathrm{Error}_{Y|X}(A,B,\boldsymbol{c}) \leq \alpha_{AB} - 1 + \frac{\beta_{AB}+1}{\kappa} + 2\kappa\left[\max\{\alpha_A, 1\}\, 2^{-n[\inf F_{Y|X}(R_A) - 2\lambda_{\mathcal{X}\mathcal{Y}}]} + \beta_A\right] \tag{10}$$

for all $\mu_{Y|X}$ satisfying

$$H(X|Y) < R_A, \tag{11}$$

where

$$F_{Y|X}(R) \equiv \min_{V|U}[D(\nu_{V|U}\|\mu_{Y|X}|\nu_U) + |R - H(U|V)|^+],$$

the infimum is taken over all $\mu_{Y|X}$ satisfying (9), and $\boldsymbol{\kappa} \equiv \{\kappa(n)\}_{n=1}^{\infty}$ is an arbitrary sequence satisfying

$$\lim_{n\to\infty} \kappa(n) = \infty \tag{12}$$

$$\lim_{n\to\infty} \kappa(n)\beta_A(n) = 0 \tag{13}$$

$$\lim_{n\to\infty} \frac{\log\kappa(n)}{n} = 0 \tag{14}$$

and κ denotes $\kappa(n)$. Since

$$\inf_{\substack{\mu_{Y|X}: \\ H(Y|X)<R_A}} F_{Y|X}(R_A) > 0,$$

then the right hand side of (10) goes to zero as $n \to \infty$ for all $\mu_{Y|X}$ satisfying (11). ■

Remark 3: It should be noted here that we have

$$I(X;Y) > R_B \tag{15}$$

from (11) and (9). However (11) and (15) do not imply (9) even when $R_A < H(X)$. ■

Remark 4: For $\boldsymbol{\beta}_A$ satisfying (H3), there is $\boldsymbol{\kappa}$ satisfying (12)–(14) by letting

$$\kappa(n) \equiv \begin{cases} n^{\xi} & \text{if } \beta_A(n) = o\left(n^{-\xi}\right) \\ \frac{1}{\sqrt{\beta_A(n)}}, & \text{otherwise} \end{cases} \tag{16}$$

for every n. If $\beta_A(n)$ is not $o\left(n^{-\xi}\right)$, there is $\kappa' > 0$ such that $\beta_A(n)n^{\xi} > \kappa'$ and

$$\begin{aligned}\frac{\log \kappa(n)}{n} &= \frac{\log \frac{1}{\beta_A(n)}}{2n} \\ &\leq \frac{\log \frac{n^{\xi}}{\kappa'}}{2n} \\ &= \frac{\xi \log n - \log \kappa'}{2n}\end{aligned}$$

for all sufficiently large n. This implies that $\boldsymbol{\kappa}$ satisfies (14). It should be noted that we can let ξ be arbitrarily large in (16) when $\beta_A(n)$ vanishes exponentially fast. This parameter ξ affects the upper bound of (10). ■

Remark 5: From Lemma 4, we have the fact that the condition (H3) of $\boldsymbol{\beta}_B$ is not necessary for the ensembles $(\mathcal{A}, p_A)$ and $(\mathcal{B}, p_B)$ of linear functions. ■

VI. Proof of Theorems

In this section, we prove the theorems.

A. *Proof of Theorem 1*

Let

$$\mathcal{G}_U \equiv \{\boldsymbol{x}' : H(\boldsymbol{x}') \leq H(U)\}.$$

If $\boldsymbol{x} \in \mathcal{T}_U$ and $g_A(A\boldsymbol{x}) \neq \boldsymbol{x}$, then there is $\boldsymbol{x}' \in \mathcal{C}_A(\mathcal{A}\boldsymbol{x})$ such that $\boldsymbol{x}' \neq \boldsymbol{x}$ and

$$H(\boldsymbol{x}') \leq H(\boldsymbol{x}) = H(U),$$

which implies that

$$[\mathcal{G}_U - \{\boldsymbol{x}\}] \cap \mathcal{C}_A(A\boldsymbol{x}) \neq \emptyset.$$

Then we have

$$\begin{aligned}E_A\left[\mathrm{Error}_X(A)\right] &= E_A\left[\sum_{\boldsymbol{x}} \mu_X(\boldsymbol{x})\chi(g_A(A\boldsymbol{x}) \neq \boldsymbol{x})\right] \\ &\leq \sum_U \sum_{\boldsymbol{x}\in\mathcal{T}_U} \mu_X(\boldsymbol{x}) p_A\left(\left\{\begin{array}{l} A : \\ [\mathcal{G}_U - \{\boldsymbol{x}\}] \cap \mathcal{C}_A(A\boldsymbol{x}) \neq \emptyset \end{array}\right\}\right) \\ &\leq \sum_U \sum_{\boldsymbol{x}\in\mathcal{T}_U} \mu_X(\boldsymbol{x}) \max\left\{\frac{|\mathcal{G}_U|\alpha_A}{|\mathrm{Im}\mathcal{A}|} + \beta_A, 1\right\} \\ &\leq \sum_U \sum_{\boldsymbol{x}\in\mathcal{T}_U} \mu_X(\boldsymbol{x}) \max\left\{\frac{|\mathcal{X}|^{l_A} 2^{-n[R-H(U)-\lambda_{\mathcal{X}}]}\alpha_A}{|\mathrm{Im}\mathcal{A}|}, 1\right\} + \beta_A \\ &\leq \max\left\{\frac{\alpha_A|\mathcal{X}|^{l_A}}{|\mathrm{Im}\mathcal{A}|}, 1\right\} \sum_U 2^{-n[D(\nu_U\|\mu_X)+|R-H(U)|^+-\lambda_{\mathcal{X}}]} + \beta_A \\ &\leq \max\left\{\frac{\alpha_A|\mathcal{X}|^{l_A}}{|\mathrm{Im}\mathcal{A}|}, 1\right\} 2^{-n[F_X(R)-2\lambda_{\mathcal{X}}]} + \beta_A,\end{aligned}$$

where the second inequality comes from Lemma 2, the third inequality comes from Lemma 8, the fourth inequality comes from Lemmas 6 and 7, and the last inequality comes from the definition of F_X and Lemma 5. Then we have the fact that there is a function (matrix) $A \in \mathcal{A}$ satisfying (5). ■

B. Proof of Theorem 3

Let $UV \equiv \nu_{VU}$ be a joint type of the sequence $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$, where the marginal type U is defined as

$$U \equiv \arg\min_{U'} D(\nu_{U'} \| \mu_X). \tag{17}$$

and the conditional type given type U is denoted by $V|U$. Since $R_A + R_B < H(X)$ and $H(U)$ approaches $H(X)$ as n goes to infinity because of the law of large numbers and the continuity of the entropy function, we have

$$H(U) - \lambda_{\mathcal{X}} > R_A + R_B + \frac{\log \kappa}{n}$$

for all sufficiently large n. Then we have

$$\begin{aligned} |\mathcal{T}_U| &\geq 2^{n[H(U)-\lambda_{\mathcal{X}}]}, \\ &\geq \kappa 2^{n[R_A+R_B]} \\ &= \kappa |\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}| \end{aligned}$$

for all sufficiently large n, where the first inequality comes from Lemma 6. This implies that there is $\mathcal{T} \subset \mathcal{T}_U$ such that

$$\kappa \leq \frac{|\mathcal{T}|}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|} \leq 2\kappa \tag{18}$$

for all sufficiently large n.

Let

$$\bullet\ g_{AB}(\boldsymbol{c},\boldsymbol{m}) \in \mathcal{T} \tag{UC1}$$

$$\bullet\ g_A(\boldsymbol{c},\boldsymbol{y}) = g_{AB}(\boldsymbol{c},\boldsymbol{m}). \tag{UC2}$$

Then we have

$$\mathrm{Error}(A, B, \boldsymbol{c}, \mu_{Y|X}) \leq p_{MY}(\mathcal{S}_1^c) + p_{MY}(\mathcal{S}_1 \cap \mathcal{S}_2^c), \tag{19}$$

where

$$\mathcal{S}_i \equiv \left\{(\boldsymbol{m},\boldsymbol{y},\boldsymbol{w}) : \text{(UC}i\text{)}\right\}.$$

First, we evaluate $E_{ABC}\left[p_{MY}(\mathcal{S}_1^c)\right]$. We have

$$\begin{aligned} E_{ABC}\left[p_{MY}(\mathcal{S}_1^c)\right] &= p_{ABCM}\left(\{(A,B,\boldsymbol{c},\boldsymbol{m}) : \mathcal{T} \cap \mathcal{C}_{AB}(\boldsymbol{c},\boldsymbol{m}) = \emptyset\}\right) \\ &\leq \alpha_{AB} - 1 + \frac{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|\left[\beta_{AB}+1\right]}{|\mathcal{T}|} \\ &\leq \alpha_{AB} - 1 + \frac{\beta_{AB}+1}{\kappa} \end{aligned} \tag{20}$$

where the equality comes from the property of $\mathcal{T}$, the first inequailty comes from Lemma 3 and the second inequality comes from (18).

Next, we evaluate $E_{ABC}\left[p_{MY}(\mathcal{S}_1 \cap \mathcal{S}_2^c)\right]$. Let

$$\mathcal{G}(\boldsymbol{y}) \equiv \{\boldsymbol{x}' : H(\boldsymbol{x}'|\boldsymbol{y}) \leq H(U|V)\}$$

and assume that $(\boldsymbol{x},\boldsymbol{y})\in\mathcal{T}_{UV}$. Then we have

$$\begin{aligned}
E_{AC}\left[\chi(A\boldsymbol{x}=\boldsymbol{c})\chi(g_A(\boldsymbol{c},\boldsymbol{y})\neq\boldsymbol{x})\right]
&= p_{AC}\left(\left\{(A,\boldsymbol{c}): \begin{array}{c} A\boldsymbol{x}=\boldsymbol{c} \\ \exists \boldsymbol{x}'\neq\boldsymbol{x} \text{ s.t.} \\ H(\boldsymbol{x}'|\boldsymbol{y})\leq H(\boldsymbol{x}|\boldsymbol{y}) \text{ and } A\boldsymbol{x}'=\boldsymbol{c}\end{array}\right\}\right) \\
&= p_A\left(\left\{A: \begin{array}{l}\exists \boldsymbol{x}'\neq\boldsymbol{x} \text{ s.t.} \\ H(\boldsymbol{x}'|\boldsymbol{y})\leq H(\boldsymbol{x}|\boldsymbol{y}) \text{ and } A\boldsymbol{x}'=A\boldsymbol{x}\end{array}\right\}\right) p_C\left(\{\boldsymbol{c}: A\boldsymbol{x}=\boldsymbol{c}\}\right) \\
&= \frac{1}{|\mathrm{Im}\mathcal{A}|} p_A\left(\left\{A: \begin{array}{l}\exists \boldsymbol{x}'\neq\boldsymbol{x} \text{ s.t. } H(\boldsymbol{x}'|\boldsymbol{y})\leq H(U|V) \\ \text{and } A\boldsymbol{x}'=A\boldsymbol{x}\end{array}\right\}\right) \\
&\leq \frac{1}{|\mathrm{Im}\mathcal{A}|}\max\left\{\sum_{\boldsymbol{x}'\in[\mathcal{G}(\boldsymbol{y})-\{\boldsymbol{x}\}]} p_A\left(\{A: A\boldsymbol{x}=A\boldsymbol{x}'\}\right), 1\right\} \\
&\leq \frac{1}{|\mathrm{Im}\mathcal{A}|}\max\left\{\frac{2^{n[H(U|V)+\lambda_{\mathcal{XY}}]}\alpha_A}{|\mathrm{Im}\mathcal{A}|}+\beta_A, 1\right\} \\
&= \frac{1}{|\mathrm{Im}\mathcal{A}|}\max\left\{2^{-n[R_A-H(U|V)-\lambda_{\mathcal{XY}}]}\alpha_A+\beta_A, 1\right\} \\
&\leq \frac{1}{|\mathrm{Im}\mathcal{A}|}\left[\max\{\alpha_A,1\}2^{-n[|R_A-H(U|V)|^+-\lambda_{\mathcal{XY}}]}+\beta_A\right],
\end{aligned} \tag{21}$$

where $|\cdot|^+$ is defined by (2), the third equality comes from Lemma 1 and the second inequality comes from Lemma 8 and (H4) for an ensemble p_A. Then we have

$$\begin{aligned}
&E_{ABC}\left[p_{MY}(\mathcal{S}_1\cap\mathcal{S}_2^c)\right] \\
&= E_{ABCM}\left[\sum_{\boldsymbol{x}\in\mathcal{T}}\sum_{V|U}\sum_{\boldsymbol{y}\in\mathcal{T}_{V|U}(\boldsymbol{x})}\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})\chi(g_{AB}(\boldsymbol{c},\boldsymbol{m})=\boldsymbol{x})\chi(g_A(\boldsymbol{c},\boldsymbol{y})\neq\boldsymbol{x})\right] \\
&\leq E_{ABCM}\left[\sum_{\boldsymbol{x}\in\mathcal{T}}\sum_{V|U}\sum_{\boldsymbol{y}\in\mathcal{T}_{V|U}(\boldsymbol{x})}\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})\chi(A\boldsymbol{x}=\boldsymbol{c})\chi(B\boldsymbol{x}=\boldsymbol{m})\chi(g_A(\boldsymbol{c},\boldsymbol{y})\neq\boldsymbol{x})\right] \\
&= \sum_{\boldsymbol{x}\in\mathcal{T}}\sum_{V|U}\sum_{\boldsymbol{y}\in\mathcal{T}_{V|U}(\boldsymbol{x})}\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})E_{AC}\left[\chi(A\boldsymbol{x}=\boldsymbol{c})\chi(g_A(\boldsymbol{c},\boldsymbol{y})\neq\boldsymbol{x})\right]E_{BM}\left[\chi(B\boldsymbol{x}=\boldsymbol{m})\right] \\
&\leq \frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\sum_{\boldsymbol{x}\in\mathcal{T}}\sum_{V|U}\sum_{\boldsymbol{y}\in\mathcal{T}_{V|U}(\boldsymbol{x})}\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})\left[\max\{\alpha_A,1\}2^{-n[|R_A-H(U|V)|^+-\lambda_{\mathcal{XY}}]}+\beta_A\right] \\
&= \frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\sum_{\boldsymbol{x}\in\mathcal{T}}\left[\sum_{V|U}\sum_{\boldsymbol{y}\in\mathcal{T}_{V|U}(\boldsymbol{x})}\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})\max\{\alpha_A,1\}2^{-n[|R_A-H(U|V)|^+-\lambda_{\mathcal{XY}}]}+\beta_A\right] \\
&\leq \frac{1}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\sum_{\boldsymbol{x}\in\mathcal{T}}\left[\max\{\alpha_A,1\}\sum_{V|U}2^{-n[D(\nu_{V|U}\|\mu_{Y|X}|\nu_U)+|R_A-H(U|V)|^+-\lambda_{\mathcal{XY}}]}+\beta_A\right] \\
&\leq \frac{|\mathcal{T}|}{|\mathrm{Im}\mathcal{A}||\mathrm{Im}\mathcal{B}|}\left[\max\{\alpha_A,1\}2^{-n[F_{Y|X}(R_A)-2\lambda_{\mathcal{XY}}]}+\beta_A\right] \\
&\leq 2\kappa\left[\max\{\alpha_A,1\}2^{-n[F_{Y|X}(R_A)-2\lambda_{\mathcal{XY}}]}+\beta_A\right],
\end{aligned} \tag{22}$$

where the second inequality comes from Lemma 1 and (21), the third inequality comes from Lemmas 7 and 6, the fourth inequality comes from the definition of $F_{Y|X}$ and Lemma 5 and the last inequality comes from

(18).

From (19), (20), and (22) we have

$$E_{ABC}\left[\mathrm{Error}_{Y|X}(A,B,\boldsymbol{c})\right] \leq \alpha_{AB} - 1 + \frac{\beta_{AB}+1}{\kappa} + 2\kappa\left[\max\{\alpha_A,1\}\,2^{-n[F_{Y|X}(R_A)-2\lambda_{\mathcal{XY}}]} + \beta_A\right].$$

Applying the above argument for all $\mu_{Y|X}$ satisfying (11) and (9), we have the fact that there are $A \in \mathcal{A}$, $B \in \mathcal{B}$, and $\boldsymbol{c} \in \mathrm{Im}A$ that satisfy (10). ■

VII. Conclusion

The fixed rate universal coding theorems are proved by using the notion of hash property. We proved the theorems of fixed-rate lossless universal source coding and fixed-rate universal channel coding. Since an ensemble of sparse matrices satisfies the hash property requirement, it is proved that we can construct universal codes by using sparse matrices.

Appendix

We introduce the following lemmas that are used in the proofs of the theorems.

Lemma 5 ([4, Lemma 2.2]): The number of different types of sequences in $\mathcal{X}^n$ is fewer than $[n+1]^{|\mathcal{X}|}$. The number of conditional types of sequences $\mathcal{X}\times\mathcal{Y}$ is fewer than $[n+1]^{|\mathcal{X}||\mathcal{Y}|}$. ■

Lemma 6 ([4, Lemma 2.3]): For a type U of a sequence in $\mathcal{X}^n$,

$$2^{n[H(U)-\lambda_{\mathcal{X}}]} \leq |\mathcal{T}_U| \leq 2^{nH(U)},$$

where $\lambda_{\mathcal{X}}$ is defined in (1). ■

Lemma 7 ([4, Lemma 2.6]):

$$\frac{1}{n}\log\frac{1}{\mu_X(\boldsymbol{x})} = H(\nu_{\boldsymbol{x}}) + D(\nu_{\boldsymbol{x}}\|\mu_X)$$
$$\frac{1}{n}\log\frac{1}{\mu_{Y|X}(\boldsymbol{y}|\boldsymbol{x})} = H(\nu_{\boldsymbol{y}|\boldsymbol{x}}|\nu_{\boldsymbol{x}}) + D(\nu_{\boldsymbol{y}|\boldsymbol{x}}\|\mu_{Y|X}|\nu_{\boldsymbol{y}}).$$

■

Lemma 8 ([11, Lemma 2]): For $\boldsymbol{y} \in \mathcal{T}_V$,

$$|\{\boldsymbol{x}' : H(\boldsymbol{x}') \leq H(U)\}| \leq 2^{n[H(U)+\lambda_{\mathcal{X}}]}$$
$$|\{\boldsymbol{x}' : H(\boldsymbol{x}'|\boldsymbol{y}) \leq H(U|V)\}| \leq 2^{n[H(U|V)+\lambda_{\mathcal{XY}}]},$$

where $\lambda_{\mathcal{X}}$ and $\lambda_{\mathcal{XY}}$ are defined by (1). ■

Proof: The first inequality of this lemma is shown by the second inequality. The second inequality is shown by

$$\begin{aligned}|\{\boldsymbol{x}' : H(\boldsymbol{x}'|\boldsymbol{y}) < H(U|V)\}| &= \sum_{\substack{U':\\ H(U'|V)\leq H(U|V)}} |\mathcal{T}_{U'|V}(\boldsymbol{y})| \\ &\leq \sum_{\substack{U':\\ H(U'|V)\leq H(U|V)}} 2^{nH(U'|V)}\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{\substack{U': \\ H(U'|V) \leq H(U|V)}} 2^{nH(U|V)} \\
&\leq [n+1]^{|\mathcal{X}||\mathcal{Y}|} 2^{nH(U|V)} \\
&= 2^{n[H(U|V)+\lambda_{\mathcal{UV}}]},
\end{aligned}
$$

where the first inequality comes from Lemma 6 and the third inequality comes from Lemma 5. ■

ACKNOWLEDGEMENTS

This paper was written while one of authors J. M. was a visiting researcher at ETH, Zürich. He wishes to thank Prof. Maurer for arranging for his stay.

REFERENCES

[1] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete-memoryless channels," *IEEE Trans. Inform. Theory*, vol. IT-50, no. 3, pp. 417–438, Mar. 2004.

[2] T. P. Coleman, M. Médard, and M. Effros, "Towards practical miminum-entropy universal decoding," *Proc. of the IEEE Data Compression Coference*, Mar. 29–31, 2005 pp. 33–42.

[3] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 4, pp. 585–592, Jul. 1982.

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, 1981.

[5] J. L. Carter and M. N. Wegman, "Universal classes of hash functions," *J. Comput. Syst. Sci.*, vol. 18, pp. 143–154, 1979.

[6] J. Feldman, M.J. Wainwright, and D.R. Karger, "Using linear programming to decode binary linear codes," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 3, pp. 954–972, Mar. 2005.

[7] H. Koga, "Source coding using families of universal hash functions," *IEEE Trans. Inform. Theory*, vol. IT-53, no. 9, pp. 3226–3233, Sept. 2007.

[8] S. Miyake and M. Maruyama, "Construction of universal codes using LDPC matrices and their error exponents," *IEICE Trans. Fundamentals*, vol. E90-A, No. 9, pp. 1830–1839, Sept. 2007.

[9] S. Miyake and J. Muramatsu, "A construction of channel code, JSCC and universal code for discrete memoryless channels using sparse matrices," to appear in *Proc. 2008 IEEE Int. Symp. Inform. Theory*, Tronto, Canada, Jul. 6–11, 2008.

[10] J. Muramatsu, T. Uyematsu, and T. Wadayama, "Low density parity check matrices for coding of correlated sources," *IEEE Trans. Inform. Theory*, vol. IT-51, no. 10, pp. 3645–3653, Oct. 2005.

[11] J. Muramatsu, "Secret key agreement from correlated source outputs using low density parity check matrices," *IEICE Trans. Fundamentals*, vol. E89-A, no. 7, pp. 2036–2046, Jul. 2006.

[12] J. Muramatsu and S. Miyake "Hash property and coding theorems for sparse matrices and maximal-likelihood coding," submittd to *IEEE Trans. Inform. Theory*, available at `arXiv:0801.3878 [cs.IT]`, 2007.